



US006147996A

United States Patent [19]

Laor et al.

[11] **Patent Number:** 6,147,996[45] **Date of Patent:** Nov. 14, 2000[54] **PIPELINED MULTIPLE ISSUE PACKET SWITCH**[75] Inventors: **Michael Laor, Sunnyvale; Martin Cieslak, Fremont, both of Calif.**[73] Assignee: **Cisco Technology, Inc., San Jose, Calif.**[21] Appl. No.: **08/511,146**[22] Filed: **Aug. 4, 1995**[51] Int. Cl.⁷ **H04J 12/56**[52] U.S. Cl. **370/394; 370/412**[58] Field of Search **370/389, 392, 370/394, 401, 412, 428, 429, 229, 230, 231, 235, 391, 465**[56] **References Cited****U.S. PATENT DOCUMENTS**

Re. 33,900	4/1992	Howson	370/105
4,437,087	3/1984	Petr	340/347 DD
4,438,511	3/1984	Baran	370/19
4,506,358	3/1985	Montgomery	370/60
4,646,287	2/1987	Larson et al.	370/60
4,677,423	6/1987	Benvenuto et al.	340/347
4,678,189	7/1987	Olson et al.	370/60
4,679,227	7/1987	Hughes-Hartogs	379/98
4,723,267	2/1988	Jones et al.	379/93
4,731,816	3/1988	Hughes-Hartogs	379/98
4,750,136	6/1988	Arpin et al.	364/514
4,757,495	7/1988	Decker et al.	370/76
4,769,810	9/1988	Eckberg, Jr. et al.	370/60
4,769,811	9/1988	Eckberg, Jr. et al.	370/60
4,827,411	5/1989	Arrowood et al.	364/300
4,833,706	5/1989	Hughes-Hartogs	379/98
4,835,737	5/1989	Herrig et al.	364/900

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

0 384 758	2/1990	European Pat. Off.	H04L 12/56
0 431 751 A1	11/1990	European Pat. Off.	H04I 12/46
WO 95/20850	8/1995	WIPO	H04L 12/56

OTHER PUBLICATIONS

Chowdhury, et al., "Alternative Bandwidth Allocation Algorithms for Packet Video in ATM Networks", 1992, IEEE Infocom 92, pp. 1061-1068.

Zhang, et al., "Rate-Controlled Static-Priority Queueing", 1993, IEEE, pp. 227-236.

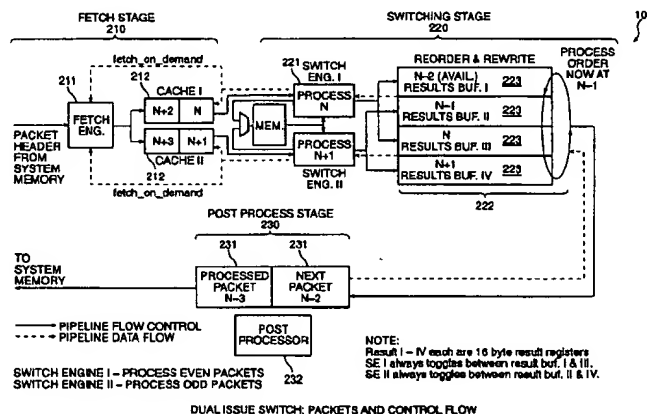
Doeringer, et al., "Routing on Longest-Matching Prefixes", IEEE ACM Transactions on Networking, Feb. 1, 1996, vol. 4, No. 1, pp. 86-97.

IBM, "Method and Apparatus for the Statistical Multiplexing of Voice, Data, and Image Signals", Nov., 1992, IBM Technical Data Bulletin n6 11-92, pp. 409-411.

(List continued on next page.)

Primary Examiner—Chi H. Pham**Assistant Examiner**—Ricky Q. Ngo**Attorney, Agent, or Firm**—Swernofsky Law Group[57] **ABSTRACT**

A pipelined multiple issue architecture for a link layer or protocol layer packet switch, which processes packets independently and asynchronously, but reorders them into their original order, thus preserving the original incoming packet order. Each stage of the pipeline waits for the immediately previous stage to complete, thus causing the packet switch to be self-throttling and thus allowing differing protocols and features to use the same architecture, even if possibly requiring differing processing times. The multiple issue pipeline is scalable to greater parallel issue of packets, and tunable to differing switch engine architectures, differing interface speeds and widths, and differing clock rates and buffer sizes. The packet switch comprises a fetch stage, which fetches the packet header into one of a plurality of fetch caches, a switching stage comprising a plurality of switch engines, each of which independently and asynchronously reads from corresponding fetch caches, makes switching decisions, and write to a reorder memory, a reorder engine which reads from the reorder memory in the packets' original order, and a post-processing stage, comprising a post-process queue and a post-process engine, which performs protocol-specific post-processing on the packets.

40 Claims, 5 Drawing Sheets

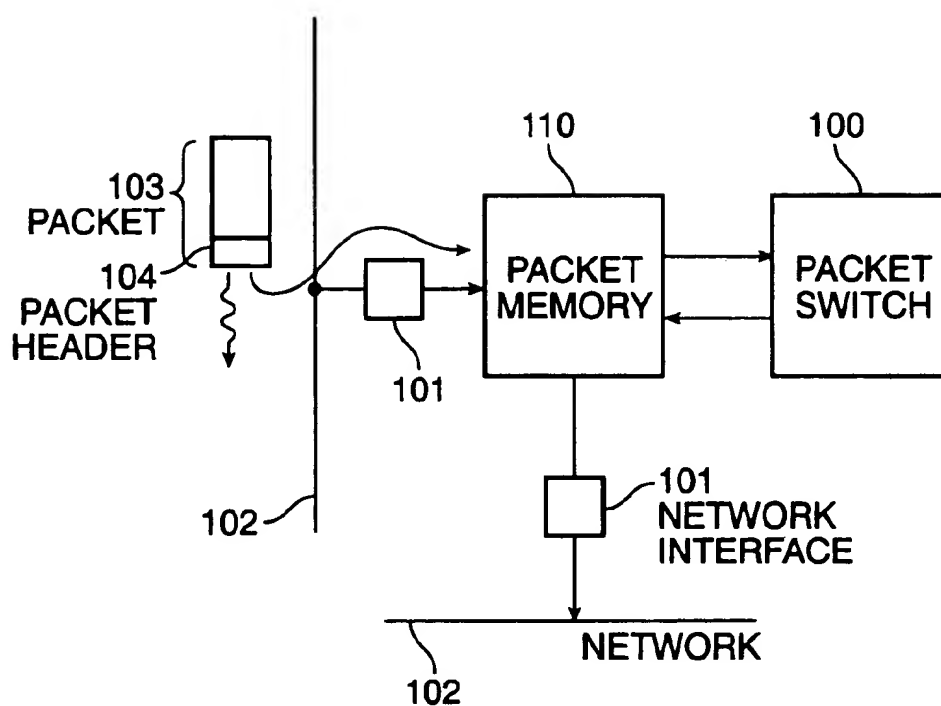
U.S. PATENT DOCUMENTS

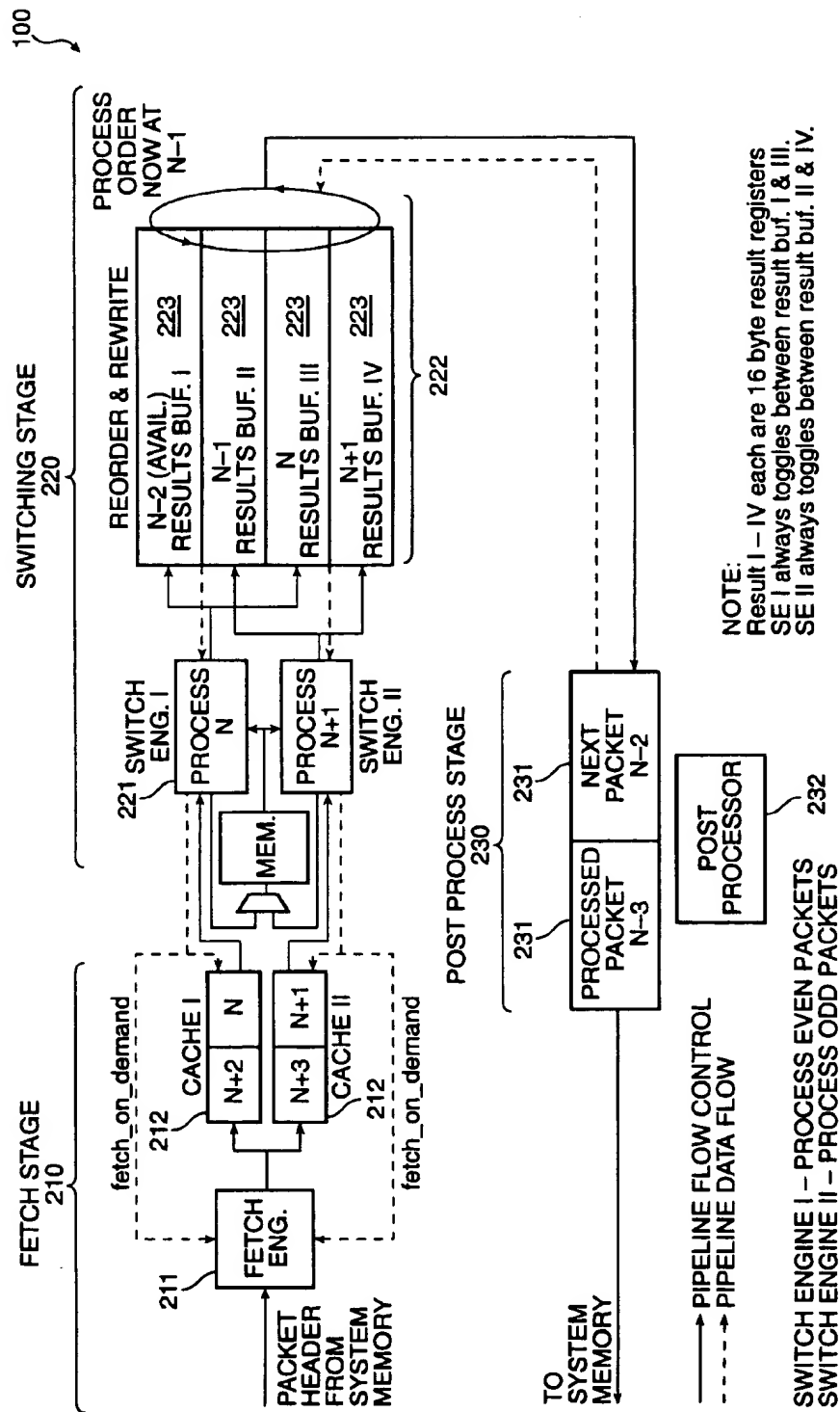
4,922,486	5/1990	Lidinsky et al.	370/60	5,546,370	8/1996	Ishikawa .	
4,960,310	10/1990	Cushing	350/1.7	5,548,593	8/1996	Peschi	370/394
4,965,772	10/1990	Daniel et al.	364/900	5,555,244	9/1996	Gupta et al.	370/60
4,979,118	12/1990	Kheradpir	364/436	5,583,862	12/1996	Callon	370/397
4,980,897	12/1990	Decker et al.	375/38	5,590,122	12/1996	Sandorfi et al.	370/394
5,014,265	5/1991	Hahne et al.	370/60	5,592,470	1/1997	Rudrapatna et al.	370/320
5,054,034	10/1991	Hughes-Hartogs	375/8	5,598,581	1/1997	Daines et al.	395/872
5,095,480	3/1992	Fenner	370/94.1	5,600,798	2/1997	Chenrukuri et al.	
5,206,889	4/1993	Bingham	375/97	5,604,868	2/1997	Komine et al.	395/200
5,208,811	5/1993	Kashio et al.		5,608,733	3/1997	Vallee et al.	370/394
5,228,062	7/1993	Bingham	375/97	5,617,417	4/1997	Sathe et al.	370/394
5,247,516	9/1993	Bernstein et al.	370/82	5,617,421	4/1997	Chin et al.	370/402
5,253,251	10/1993	Aramaki	370/394	5,630,125	5/1997	Zellweger	395/614
5,280,470	1/1994	Buhrke et al.	370/13	5,631,908	5/1997	Saxe .	
5,287,103	2/1994	Kasprzyk et al.	340/825.52	5,632,021	5/1997	Jennings et al.	395/309
5,287,453	2/1994	Roberts	395/200	5,634,010	5/1997	Ciscon et al.	395/200
5,309,437	5/1994	Perlman et al.	730/85.13	5,644,718	7/1997	Belove et al.	395/200
5,327,421	7/1994	Hiller et al.	370/61.1	5,666,353	9/1997	Klausmeier et al.	370/230
5,345,445	9/1994	Hiller et al.	370/60.1	5,673,265	9/1997	Gupta et al.	370/432
5,345,446	9/1994	Hiller et al.	370/60.1	5,678,006	10/1997	Valizadeh et al.	395/200
5,365,524	11/1994	Hiller et al.	370/94.2	5,680,116	10/1997	Hashimoto et al.	
5,367,517	11/1994	Cidon et al.	370/54	5,684,797	11/1997	Aznar et al.	370/390
5,371,852	12/1994	Attanasio et al.	395/200	5,687,324	11/1997	Green et al.	
5,390,175	2/1995	Hiller et al.	370/60	5,689,506	11/1997	Chiussi et al.	370/388
5,414,705	5/1995	Therasse et al.	370/394	5,724,351	3/1998	Chao et al.	
5,422,882	6/1995	Hiller et al.	370/60.1	5,748,186	5/1998	Raman	345/302
5,426,636	6/1995	Hiller et al.	370/60.1	5,754,547	5/1998	Nakazawa .	
5,428,607	6/1995	Hiller et al.	370/60.1	5,802,054	9/1998	Bellenger .	
5,430,729	7/1995	Rahnema .		5,835,710	11/1998	Nagami et al.	
5,442,457	8/1995	Najafi	385/400	5,854,903	12/1998	Morrison et al.	
5,452,297	9/1995	Hiller et al.	370/60.1	5,856,981	1/1999	Voelker .	
5,477,541	12/1995	White et al.		5,892,924	4/1999	Lyon et al.	395/200.75
5,483,523	1/1996	Nederlof	370/394	5,898,686	4/1999	Virgile .	
5,485,455	1/1996	Dobbins et al.	370/60	5,903,559	5/1999	Acharya et al.	
5,490,140	2/1996	Abensour et al.					
5,490,258	2/1996	Fenner	395/401				
5,497,371	3/1996	Ellis et al.	370/394				
5,519,858	5/1996	Walton et al.	395/600				
5,535,195	7/1996	Lee	370/54				
5,539,734	7/1996	Burwell et al.					
5,541,911	7/1996	Nilakantan et al.					

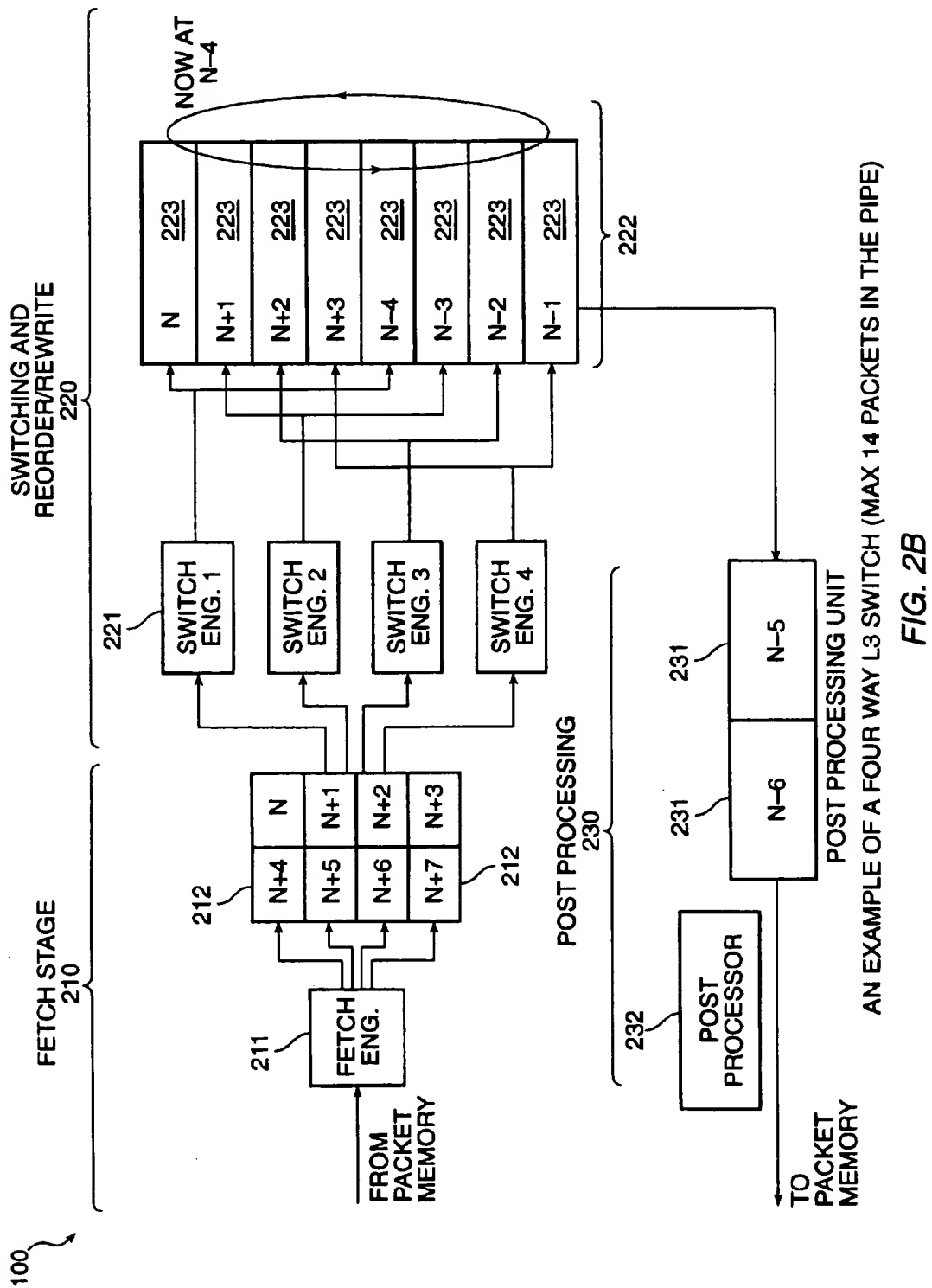
OTHER PUBLICATIONS

Esaki, et al., "Datagram Delivery in an ATM-Internet," IEICE Transactions on Communications vol. E77-B, No. 3, (1994) Mar. Tokyo, Japan.

William Stallings, Data and Computer Communications, pp. 329-333, Prentice Hall, Upper Saddle River, New Jersey 07458.

**FIG. 1**





AN EXAMPLE OF A FOUR WAY L3 SWITCH (MAX 14 PACKETS IN THE PIPE)

FIG. 2B

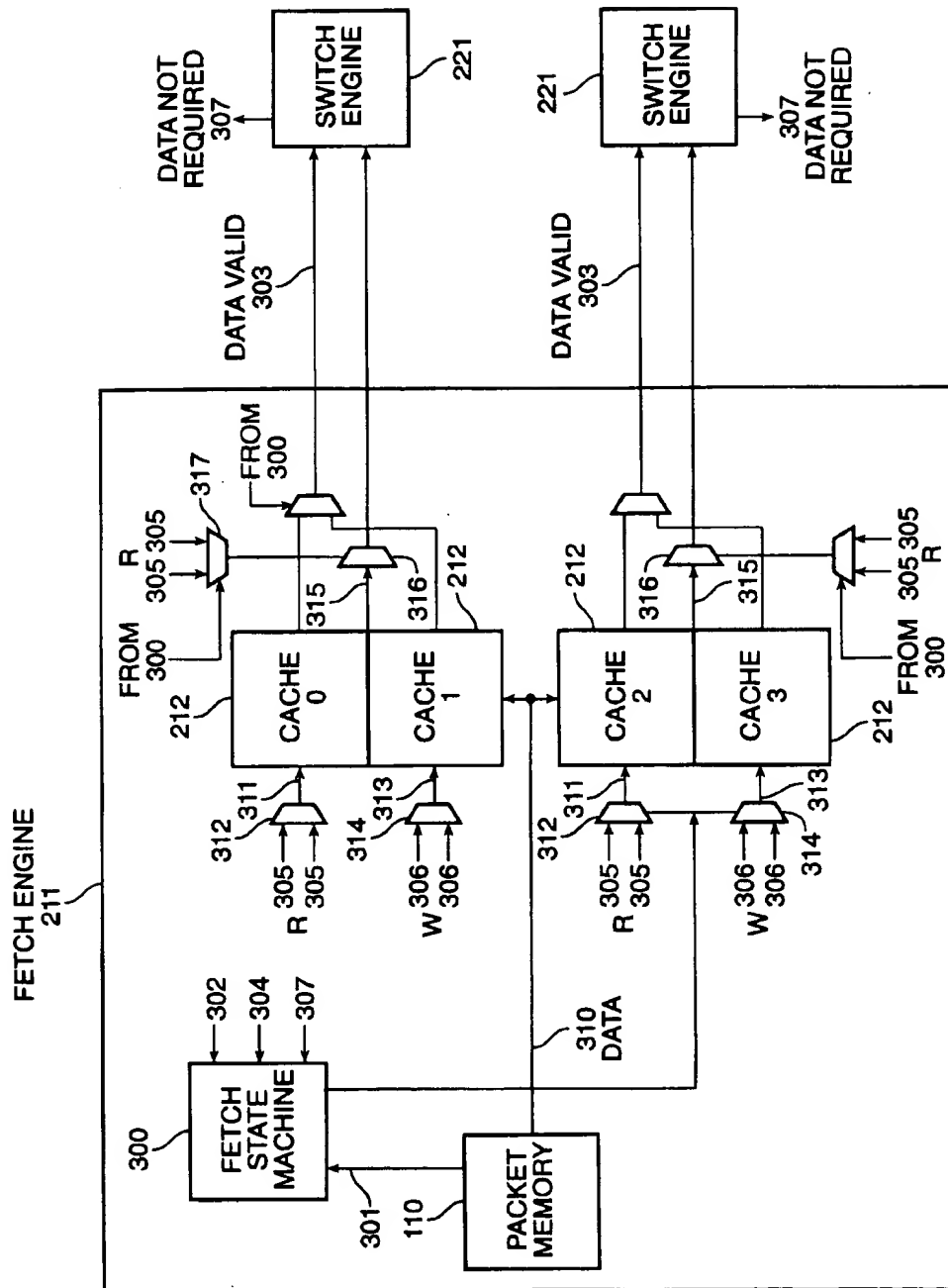


FIG. 3

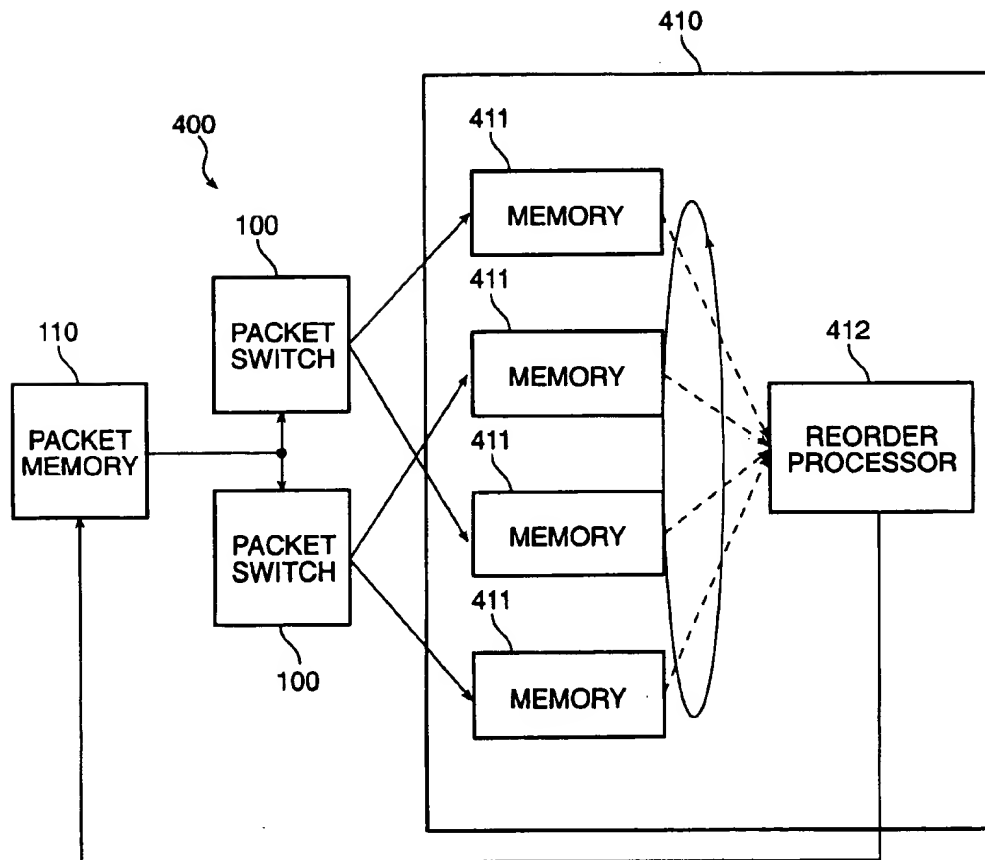


FIG. 4

PIPELINED MULTIPLE ISSUE PACKET SWITCH

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to a pipelined multiple issue packet switch.

2. Description of Related Art

When computers are coupled together into networks for communication, it is known to couple networks together and to provide a switching device which is coupled to more than one network. The switching device receives packets from one network and retransmits those packets (possibly in another format) on another network. In general, it is desirable for the switching device to operate as quickly as possible.

However, there are several constraints under which the switching device must operate. First, packets may encapsulate differing protocols, and thus may differ significantly in length and in processing time. Second, when switching packets from one network to another, it is generally required that packets are re-transmitted in the same order as they arrive. Because of these two constraints, known switching device architectures are not able to take advantage of significant parallelism in switching packets.

It is also desirable to account ahead of time for future improvements in processing hardware, such as bandwidth and speed of a network interface, clock speed of a switching processor, and memory size of a packet buffer, so that the design of the switching device is flexible and scaleable with such improvements.

The following U.S. Patents may be pertinent:

U.S. Pat. No. 4,446,555 to Devault et al., "Time Division Multiplex Switching Network For Multiservice Digital Networks";

U.S. Pat. No. 5,212,686 to Joy et al., "Asynchronous Time Division Switching Arrangement and A Method of Operating Same";

U.S. Pat. No. 5,271,004 to Proctor et al., "Asynchronous Transfer Mode Switching Arrangement Providing Broadcast Transmission"; and

U.S. Pat. No. 5,307,343 to Bostica et al., "Basic Element for the Connection Network of A Fast Packet Switching Node".

Accordingly, it would be advantageous to provide an improved architecture for a packet switch, which can make packet switching decisions responsive to link layer (ISO level 2) or protocol layer (ISO level 3) header information, which is capable of high speed operation at relatively low cost, and which is flexible and scaleable with future improvements in processing hardware.

SUMMARY OF THE INVENTION

The invention provides a pipelined multiple issue link layer or protocol layer packet switch, which processes packets independently and asynchronously, but reorders them into their original order, thus preserving the original incoming packet order. Each stage of the pipeline waits for the immediately previous stage to complete, thus causing the packet switch to be self-throttling and thus allowing differing protocols and features to use the same architecture, even if possibly requiring differing processing times. The multiple issue pipeline is scaleable to greater parallel issue of packets, and tunable to differing switch engine architectures, differing interface speeds and widths, and differing clock rates and buffer sizes.

In a preferred embodiment, the packet switch comprises a fetch stage, which fetches the packet header into one of a plurality of fetch caches, a switching stage comprising a plurality of switch engines, each of which independently and asynchronously reads from corresponding fetch caches, makes switching decisions, and writes to a reorder memory, a reorder engine which reads from the reorder memory in the packets' original order, and a post-processing stage, comprising a post-process queue and a post-process engine, which performs protocol-specific post-processing on the packets.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the placement of a packet switch in an internetwork.

FIG. 2 shows a block diagram of a packet switch. FIG. 2 comprises FIG. 2A and FIG. 2B collectively.

FIG. 3 shows a fetch stage for the packet switch.

FIG. 4 shows a block diagram of a system having a plurality of packet switches in parallel.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following description, a preferred embodiment of the invention is described with regard to preferred process steps, data structures, and switching techniques. However, those skilled in the art would recognize, after perusal of this application, that embodiments of the invention may be implemented using a set of general purpose computers operating under program control, and that modification of a set of general purpose computers to implement the process steps and data structures described herein would not require undue invention.

The present invention may be used in conjunction with technology disclosed in the following copending application.

Application Ser. No. 08/229,289, filed Apr. 18, 1994, in the name of inventors Bruce A. Wilford, Bruce Sherry, David Tsiang, and Anthony Li, titled "Packet Switching Engine" now U.S. Pat. No. 5,509,006.

This application is hereby incorporated by reference as if fully set forth herein, and is referred to herein as the "Packet Switching Engine disclosure".

PIPELINED, MULTIPLE ISSUE PACKET SWITCH

FIG. 1 shows the placement of a packet switch in an internetwork.

A packet switch 100 is coupled to a first network interface 101 coupled to a first network 102 and a second network interface 101 coupled to a second network 102. When a packet 103 is recognized by the first network interface 101 (i.e., the MAC address of the packet 103 is addressed to the packet switch 100 or to an address known to be off the first network 102), the packet 103 is stored in a packet memory 110 and a pointer to a packet header 104 for the packet 103 is generated.

In a preferred embodiment, the packet header 104 comprises a link layer (level 2) header, and a protocol layer (level 3) header. The link layer header, sometimes called a "MAC" (media access control) header, comprises information for communicating the packet 103 on a network 102 using particular media, such as the first network 102. The protocol layer header comprises information for level 3

switching of the packet 103 among networks 102. The link layer header comprises information for level 2 switching (i.e., bridging). For example, the link layer header may comprise an ethernet, FDDI, or token ring header, while the protocol layer header may comprise an IP header. Also, there are hybrid switching techniques which respond to the both the level 2 and the level 3 headers, as well as those which respond to level 4 headers (such as extended access lists). Those skilled in the art will recognize, after perusal of this application, that other types of packet headers or trailers are within the scope and spirit of the invention, and that adapting the invention to switching such packet headers would not involve invention or undue experimentation.

The packet switch 100 reads the packet header 104 and performs two tasks—(1) it rewrites the packet header 104, if necessary, to conform to protocol rules for switching the packet 103, and (2) it queues the packet 103 for transmission on an output network interface 101 and thus an output network 102. For example, if the output network 102 requires a new link layer header, the packet switch 100 rewrites the link layer header. If the protocol layer header comprises a count of the number of times the packet 103 has been switched, the packet switch 100 increments or decrements that count, as appropriate, in the protocol layer header.

FIG. 2 shows a block diagram of a packet switch. FIG. 2 comprises FIG. 2A and FIG. 2B collectively.

The packet switch 100 comprises a fetch stage 210, a switching stage 220, and a post-processing stage 230.

The pointer to the packet header 104 is coupled to the fetch stage 210. The fetch stage 210 comprises a fetch engine 211 and a plurality of (preferably two) fetch caches 212. Each fetch cache 212 comprises a double buffered FIFO queue.

FIG. 2A shows a preferred embodiment in which there are two fetch caches 212, while FIG. 2B shows an alternative preferred embodiment in which there are four fetch caches 212.

In response to a signal from the switching stage 220, the fetch engine 211 prefetches a block of M bytes of the packet header 104 and stores that block in a selected FIFO queue of a selected fetch cache 212. In a preferred embodiment, the value of M, the size of the block, is independent of the protocol embodied in the protocol link layer, and is preferably about 64 bytes. In alternative embodiments, the value of M may be adjusted, e.g., by software, so that the packet switch 100 operates most efficiently with a selected mix of packets 103 it is expected to switch.

When the block of M bytes does not include the entire packet header 104, the fetch engine 211 fetches, in response to a signal from the fetch cache 212, a successive block of L additional bytes of the packet header 104 and stores those blocks in the selected FIFO queue of the selected fetch cache 212, thus increasing the amount of data presented to the switching stage 220. In a preferred embodiment, the value of L, the size of the additional blocks, is equal to the byte width of an interface to the packet memory 110, and in a preferred embodiment is about 8 bytes.

After storing at least a portion of a packet header 104 in a fetch cache 212, the fetch engine 211 reads the next packet header 104 and proceeds to read that packet header 104 and store it in a next selected fetch cache 212. The fetch caches 212 are selected for storage in a round-robin manner. Thus when there are N fetch caches 212, each particular fetch cache 212 receives every Nth packet header 104 for storage; when there are two fetch caches 212, each particular fetch cache 212 receives every other packet header 104 for storage.

Each fetch cache 212 is double buffered, so that the fetch engine 211 may write a new packet header 104 to a fetch cache 212 while the corresponding switch engine 221 is reading from the fetch cache 212. This is in addition to the fetch on demand operation described above, in which the fetch engine 211 writing successive blocks of additional bytes of an incomplete packet header 104 in response to a signal from a switch engine 221. Thus each particular fetch cache 212 pipelines up to two packet headers 104; when there are N fetch caches 212, there are up to 2N packet headers 104 pipelined in the fetch stage 210.

More generally, there may be N fetch caches 212, each of which comprises B buffers, for a total of BN buffers. The fetch engine 211 writes new packet headers 104 in sequence to the N fetch caches 212 in order, and when the fetch engine 211 returns to a fetch cache 212 after writing in sequence to all other fetch caches 212, it writes in sequence to the next one of the B buffers within that fetch cache 212.

As shown below, the switching stage 220 comprises an identical number N of switch engines 221, each of which reads in sequence from one of the B buffers of its designated fetch cache 212, returning to read from a buffer after reading in sequence from all other buffers in that fetch cache 212.

In FIG. 2A, a preferred embodiment in which there are two fetch caches 212, there are four packet headers 104 pipelined in the fetch stage 210, labeled n+3, n+2, n+1, and n. In FIG. 2B, an alternative preferred embodiment in which there are four fetch caches 212, there are eight packet headers 104 pipelined in the fetch stage 210, labeled n+7, n+6, n+5, n+4, n+3, n+2, n+1, and n.

The fetch stage 210 is further described with regard to FIG. 3.

The switching stage 220 comprises a plurality of switch engines 221, one for each fetch cache 212, and a reorder/rewrite engine 222.

Each switch engine 221 is coupled to a corresponding fetch cache 212. Each switch engine 221 independently and asynchronously reads from its corresponding fetch cache 212, makes a switching decision, and writes its results to one of a plurality of (preferably two) reorder/rewrite memories 223 in the reorder/rewrite engine 222. Thus, when there are N fetch caches 212, there are also N switch engines 221, and when there are K reorder/rewrite memories 223 for each switch engine 221, there are KN reorder/rewrite memories 223 in N sets of K.

FIG. 2A shows a preferred embodiment in which there are two switch engines 221 and four reorder/rewrite memories 223, while FIG. 2B shows an alternative preferred embodiment in which there are four switch engines 221 and eight reorder/rewrite memories 223.

In a preferred embodiment, each switch engine 221 comprises a packet switch engine as shown in the Packet Switching Engine disclosure. The switching results and other data (e.g., statistical information) written into the reorder/rewrite memories 223 comprise information regarding how to rewrite the packet header 104 and to which network interface 101 to output the packet 103. Preferably, this information comprises results registers as described in the Packet Switching Engine disclosure, and includes a pointer to the packet header 104 in the packet memory 110.

Preferably, a single integrated circuit chip comprises significant circuits of at least one, and preferably more than one, switch engine 221.

As described in the Packet Switching Engine disclosure, each switch engine 221 reads instructions from a "tree

memory" comprising instructions for reading and interpreting successive bytes of the packet header 104. In a preferred embodiment, the tree memory comprises a set of memory registers coupled to the switch engine 221. In an alternative embodiment, at least some of the tree memory may be cached on the integrated circuit chip for the switch engine 221.

The reorder/rewrite engine 222 reads from the reorder/rewrite memories 223 in a preselected order. The N sets of K reorder/rewrite memories 223 are interleaved, so that results from the switch engines 221 are read in a round-robin manner. Thus, output from the reorder/rewrite engine 222 is in the original order in which packets 103 arrived at the packet switch 100.

Thus, each one of the switch engines 221 writes in sequence to its K designated reorder/rewrite memories 223, returning to one of its designated reorder/rewrite memories 223 after writing in sequence to its other designated reorder/rewrite memories 223. In parallel, the reorder/rewrite engine 222 reads in sequence from all the NK reorder/rewrite memories 223, and returns to one of the NK reorder/rewrite memories 223 after reading in sequence from all other reorder/rewrite memories 223.

In FIG. 2A, a preferred embodiment in which there are two switch engines 221 and four reorder/rewrite memories 223, there are four packet headers 104 pipelined in the switching stage 220, labeled n+1, n, n-1, and n-2 (now available). In FIG. 2B, an alternative preferred embodiment in which there are four switch engines 221 and eight reorder/rewrite memories 223, there are eight packet headers 104 pipelined in the switching stage 220, labeled n+3, n+2, n+1, n, n-1, n-2, n-3, and n-4.

The reorder/rewrite engine 222, in addition to receiving the packet headers 104 in their original order from the reorder/rewrite memories 223, may also rewrite MAC headers for the packet headers 104 in the packet memory 110, if such rewrite is called for by the switching protocol.

The post-processing stage 230 comprises a post-processing queue 231 and a post-processor 232.

The reorder/rewrite engine 222 writes the packet headers 104 into a FIFO queue of post-processing memories 231 in the order it reads them from the reorder/rewrite memories 223. Because the queue is a FIFO queue, packet headers 104 leave the post-processing stage 230 in the same order they enter, which is the original order in which packets 103 arrived at the packet switch 100.

The post-processor 232 performs protocol-specific operations on the packet header 104. For example, the post-processor 232 increments hop counts and recomputes header check-sums for IP packet headers 104. The post-processor 232 then queues the packet 103 for the designated output network interface 101, or, if the packet 103 cannot be switched, discards the packet 103 or queues it for processing by a route server, if one exists.

In FIG. 2A, a preferred embodiment, and in FIG. 2B, an alternative preferred embodiment, there are two post-processing memories 231 in the FIFO queue for the post-processing stage 230. In FIG. 2A there are two packet headers 104 pipelined in the post-processing stage 230, labeled n-3 and n-2. In FIG. 2B there are two packet headers 104 pipelined in the post-processing stage 230, labeled n-6 and n-5.

FIG. 2A, a preferred embodiment, and FIG. 2B, an alternative preferred embodiment, show that there are several packet headers 104 processed in parallel by the packet switch 100. In general, where there are S switching engines

211, there are $3S+2$ packet headers 104 processed in parallel by the packet switch 100. Of these, 2S packet headers 104 are stored in the fetch stage 210, S packet headers 104 are stored in the reorder/rewrite memories 223, and 2 packet headers 104 are stored in the post-processing stage 230.

In a preferred embodiment, the packet memory 110 is clocked at about 50 MHz and has a memory fetch path to the fetch stage 210 which is eight bytes wide, there are two switching engines 221, each of which operates at an average switching speed of about 250 kilopackets switched per second, and each stage of the packet switch 100 completes operation within about 2 microseconds. Although each switching engine 221 is individually only about half as fast as the pipeline processing speed, the accumulated effect when using a plurality of switching engines 221 is to add their effect, producing an average switching speed for the packet switch 100 of about 500 kilopackets switched per second when the pipeline is balanced.

In an alternative preferred embodiment, each switching engine 221 operates at an average switching speed of about 125 kilopackets switched per second, producing an average switching speed for the packet switch 100 of about 250 kilopackets switched per second when the pipeline is balanced. Because the pipeline is limited by its slowest stage, the overall speed of the packet switch 100 is tunable by adjustment of parameters for its architecture, including speed of the memory, width of the memory fetch path, size of the cache buffers, and other variables. Such tunability allows the packet switch 100 to achieve satisfactory performance at a reduced cost.

FETCH ENGINE AND FETCH MEMORIES

FIG. 3 shows a fetch stage for the packet switch.

The fetch engine 211 comprises a state machine 300 having signal inputs coupled to the packet memory 110 and to the switching stage 220, and having signal outputs coupled to the switching stage 220.

A packet ready signal 301 is coupled to the fetch engine 211 from the packet memory 110 and indicates whether there is a packet header 104 ready to be fetched. In this description of the fetch engine 211, it is presumed that packets 103 arrive quickly enough that the packet ready signal 301 indicates that there is a packet header 104 ready to be fetched at substantially all times. If the fetch engine 211 fetches packet headers 104 quicker than those packet headers 104 arrive, at some times the fetch engine 211 (and the downstream elements of the packet switch 100) will have to wait for more packets 103 to switch.

A switch ready signal 302 is coupled to the fetch engine 211 from each of the switch engines 221 and indicates whether the switch engine 211 is ready to receive a new packet header 104 for switching.

A data available (or cache ready) signal 303 is coupled to each of the switch engines 221 from the fetch engine 211 and indicates whether a packet header 104 is present in the fetch cache 212 for switching.

A cache empty signal 304 is coupled to the fetch engine 211 from each of the fetch caches 212 and indicates whether the corresponding switch engine 211 has read all the data from the packet header 104 supplied by the fetch engine 211. A data not required signal 307 is coupled to the fetch engine 211 from each of the switch engines 211 and indicates whether the switch engine 211 needs further data loaded into the fetch cache 212.

It may occur that the switch engine 211 is able to make its switching decision without need for further data from the

packet header 104, even though the switch engine 211 has read all the data from the packet header 104 supplied by the fetch engine 211. In this event, the switch engine 211 sets the data not required signal 307 to inform the fetch engine 211 that no further data should be supplied, even though the cache empty signal 304 has been triggered.

It may also occur that the switch engine 211 is able to determine that it can make its switching decision within the data already available, even if it has not made that switching decision yet. For example, in the IP protocol, it is generally possible to make the switching decision with reference only to the first 64 bytes of the packet header 104. If the switch engine 211 is able to determine that a packet header 104 is an IP packet header, it can set the data not required signal 307.

A read pointer 305 is coupled to each of the fetch caches 212 from the corresponding switch engine 221 and indicates a location in the fetch cache 212 where the switch engine 221 is about to read a word (of a packet header 104) from the fetch cache 212.

A write pointer 306 is coupled to each of the fetch caches 212 from the fetch engine 211 and indicates a location in the fetch cache 212 where the fetch engine 211 is about to write a word (of a packet header 104) to the fetch cache 212.

A first pair of fetch caches 212 (labeled "0" and "1") and a second pair of fetch caches 212 (labeled "2" and "3") each comprise dual port random access memory (RAM), preferably a pair of 16 word long by 32 bit wide dual port RAM circuits disposed to respond to addresses as a single 16 word long by 64 bit wide dual port RAM circuit.

A 64 bit wide data bus 310 is coupled to a data input for each of the fetch caches 212.

The read pointers 305 for the first pair of the fetch caches 212 (labeled as "0" and "1") are coupled to a first read address bus 311 for the fetch caches 212 using a first read address multiplexer 312. The two read pointers 305 are data inputs to the read address multiplexer 312; a select input to the read address multiplexer 312 is coupled to an output of the fetch engine 211. Similarly, the read pointers 305 for the second pair of the fetch caches 212 (labeled as "2" and "3") are coupled to a second read address bus 311 for the fetch caches 212 using a second read address multiplexer 312, and selected by an output of the fetch engine 211.

Similarly, the write pointers 306 for the first pair of the fetch caches 212 (labeled as "0" and "1") are coupled to a first write address bus 313 for the fetch caches 212 using a first write address multiplexer 314. The two write pointers 306 are data inputs to the write address multiplexer 314; a select input to the write address multiplexer 314 is coupled to an output of the fetch engine 211. Similarly, the write pointers 306 for the second pair of the fetch caches 212 (labeled as "2" and "3") are coupled to a second write address bus 313 for the fetch caches 212 using a second write address multiplexer 314, and selected by an output of the fetch engine 211.

An output 315 for the first pair of fetch caches 212 is coupled to a byte multiplexer 316. The byte multiplexer 316 selects one of eight bytes of output data, and is selected by an output of a byte select multiplexer 317. The byte select multiplexer 317 is coupled to a byte address (the three least significant bits of the read pointer 305) for each of the first pair of fetch caches 212, and is selected by an output of the fetch engine 211.

An initial value for the byte address (the three least significant bits of the read pointer 305) may be set by the state machine 300 to allow a first byte of the packet header

104 to be offset from (i.e., not aligned with) an eight-byte block in the packet memory 110. The state machine 300 resets the byte address to zero for successive sets of eight bytes to be fetched from the packet memory 110.

Similarly, an output 315 for the second pair of fetch caches 212 is coupled to a byte multiplexer 316. The byte multiplexer 316 selects one of eight bytes of output data, and is selected by an output of a byte select multiplexer 317. The byte select multiplexer 317 is coupled to a byte address (the three least significant bits of the read pointer 305) for each of the second pair of fetch caches 212, and is selected by an output of the fetch engine 211. The byte multiplexers 316 are coupled to the switching stage 220.

As described with regard to FIG. 2, the fetch engine 211 responds to the switch ready signal 302 from a switch engine 221 by prefetching the first M bytes of the packet header 104 from the packet memory 110 into the corresponding fetch cache 212. To perform this task, the fetch engine 211 selects the write pointer 306 for the corresponding fetch cache 212 using the corresponding write address multiplexer 314, writes M bytes into the corresponding fetch cache 212, and updates the write pointer 306.

As described with regard to FIG. 2, the fetch cache 212 raises the cache empty signal 304 when the read pointer 305 approaches the write pointer 306, such as when the read pointer 305 is within eight bytes of the write pointer 306. The fetch engine 211 responds to the cache empty signal 304 by fetching the next L bytes of the packet header 104 from the packet memory 110 into the corresponding fetch cache 212, unless disabled by the data not required signal 307 from the switch engine 221. To perform this task, the fetch engine 211 proceeds in like manner as when it prefetched the first M bytes of the packet header 104.

In a preferred embodiment, the fetch cache 212 includes a "watermark" register (not shown) which records an address value which indicates, when the read pointer 305 reaches that address value, that more data should be fetched. For example, the watermark register may record a value just eight bytes before the write pointer 306, so that more data will only be fetched when the switch engine 221 is actually out of data, or the watermark register may record a value more bytes before the write pointer 306, so that more data will be fetched ahead of actual need. Too-early values may result in data being fetched ahead of time without need, while too-late values may result in the switch engine 221 having to wait. Accordingly, the value recorded in the watermark register can be adjusted to better match the rate at which data is fetched to the rate at which data is used by the switch engine 221.

While the switch engine 221 reads from the fetch cache 212, the fetch engine 211 prefetches the first M bytes of another packet header 104 from the packet memory 110 into another fetch cache 212 (which may eventually comprise the other fetch cache 212 of the pair). To perform this task, the fetch engine 211 selects the write pointer 306 for the recipient fetch cache 212 using the corresponding write address multiplexer 314, writes M bytes into the recipient fetch cache 212, and updates the corresponding write pointer 306.

The switch engines 221 are each coupled to the read pointer 305 for their corresponding fetch cache 212. Each switch engine 221 independently and asynchronously reads from its corresponding fetch cache 212 and processes the packet header 104 therein. To perform this task, the switch engine 221 reads one byte at a time from the output of the output multiplexer 320 and updates the corresponding byte

address (the three least significant bits of the read pointer 305). When the read pointer 305 approaches the write pointer 306, the cache low signal 304 is raised and the fetch engine 211 fetches L additional bytes "on demand".

MULTIPLE PACKET SWITCHES IN PARALLEL

FIG. 4 shows a block diagram of a system having a plurality of packet switches in parallel.

In a parallel system 400, the packet memory 110 is coupled in parallel to a plurality of (preferably two) packet switches 100, each constructed substantially as described with regard to FIG. 1. Each packet switch 100 takes its input from the packet memory 110. However, the output of each packet switch 100 is directed instead to a reorder stage 410, and an output of the reorder stage 410 is directed to the packet memory 110 for output to a network interface 101.

The output of each packet switch 100 is coupled in parallel to the reorder stage 410. The reorder stage 410 comprises a plurality of reorder memories 411, preferably two per packet switch 100 for a total of four reorder memories 411. The reorder stage 410 operates similarly to the reorder/rewrite memories 222 of the packet switch 100; the packet switches 100 write their results to the reorder memories 411, whereinafter a reorder processor 412 reads their results from the reorder memories 411 and writes them in the original arrival order of the packets 103 to the packet memory 110 for output to a network interface 101.

In a preferred embodiment where each packet switch 100 operates quickly enough to achieve an average switching speed of about 500 kilopackets per second and the reorder stage 410 operates quickly enough so that the pipeline is still balanced, the parallel system 400 produces a throughput of about 1,000 kilopackets switched per second.

Alternative embodiments of the parallel system 400 may comprise larger numbers of packet switches 100 and reorder/rewrite memories 411. For example, in one alternative embodiment, there are four packet switches 100 and eight reorder/rewrite memories 411, and the reorder stage 410 is greatly speeded up. In this alternative embodiment, the parallel system 400 produces a throughput of about 2,000 kilopackets switched per second.

Alternative Embodiments

Although preferred embodiments are disclosed herein, many variations are possible which remain within the concept, scope, and spirit of the invention, and these variations would become clear to those skilled in the art after perusal of this application.

We claim:

1. A packet switch comprising

a fetch engine coupled to a source of packet headers;
a plurality of fetch caches coupled to said fetch engine, and disposed to store at least portions of packet headers received therefrom;

a plurality of switch engines, each coupled to a corresponding one of said fetch caches, and disposed to read said portions of packet headers therefrom;

a plurality of reorder/rewrite buffers, each said reorder/rewrite buffer coupled to one of said switch engines, and disposed to store pointers to packet headers received from at least one of said plurality of switch engines;

a reorder/rewrite engine coupled to said plurality of reorder/rewrite buffers, and disposed to read pointers to packet headers therefrom in an order said packet headers were originally received;

a post-process queue coupled to said reorder/rewrite engine, and disposed to store pointers to packet headers received therefrom; and

a post-process engine coupled to said post-process queue, and disposed to process said packet headers.

2. A packet switch comprising:

a switching stage;

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to said switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post process stage; and

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers.

3. A packet switch as in claim 2, wherein said switching stage comprises a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results for switching said packet header such that the input order of said packet header is preserved.

4. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage, wherein said fetch stage comprises a fetch engine, said fetch engine being disposed to fetch a first block of M bytes of a packet header in response to a first signal, and being disposed to fetch an additional block of L bytes of said packet header in response to a second signal;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage; and

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers.

5. A packet switch as in claim 4, wherein said first signal indicates an empty fetch cache and wherein said second signal indicates a fetch cache with fewer than a selected number of unread bytes.

6. A packet switch as in claim 4, wherein M is independent of said packet header.

7. A packet switch as in claim 4, wherein M is adjusted according to said packet header.

8. A packet switch as in claim 4, wherein L is equal to a byte width of an interface to said source of packet headers.

9. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage, wherein said fetch stage comprises a plurality of fetch caches, each one of said fetch caches being coupled to said source of packet headers and each being disposed to store at least a portion of a packet header;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage; and

11

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers.

10. A packet switch as in claim 9, wherein said switching stage comprises a plurality of switch engines, each being coupled to one said fetch cache and each being disposed to receive said portion of a packet header and to produce a set of results for switching said packet header.

11. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage, wherein said fetch stage comprises a fetch engine coupled to said source of packet headers;

a plurality of fetch caches coupled to said fetch engine, each said fetch cache comprising a plurality of buffers;

wherein said fetch engine is disposed to write at least a portion of each said packet header in sequence to each said fetch cache in a selected buffer thereof;

wherein said switching stage comprises a switch engine for each said fetch cache, wherein each said switch engine is disposed to read at least a portion of said packet header in sequence from each said buffer of said fetch cache;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage; and

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers.

12. A packet switch as in claim 11, wherein each said fetch cache is selected for storage in a round-robin manner.

13. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage, wherein said switching stage comprises a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a pointer to a packet header;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage; and

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers.

14. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage;

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers; and

wherein said switching stage comprises

12

a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results for switching said packet header;

a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a packet header; and

a reorder/rewrite processor coupled to said plurality of reorder/rewrite memories and disposed to receive said packet headers from said reorder/rewrite memories in an order in which said packet headers were originally received.

15. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage;

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers; and

wherein said switching stage comprises

a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results for switching said packet header; and

a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a packet header;

said reorder/rewrite memories being divided into sets, each said set of reorder/rewrite memories being assigned to and receiving outputs from exactly one said switch engine.

16. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage;

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers; and

wherein said switching stage comprises

a plurality of switching engines, each said switching engine having a plurality of reorder/rewrite memories coupled thereto, each said switching engine being disposed to write in sequence to one of said plurality of reorder/rewrite memories; and

a reorder/rewrite engine coupled to all said reorder/rewrite memories, said reorder/rewrite engine being disposed to read in sequence from said reorder/rewrite memories.

17. A packet switch as in claim 16, wherein said reorder/rewrite engine is disposed to alter at least portions of packet headers referenced by said reorder/rewrite memories.

18. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage;

13

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage;

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers, wherein said post-process stage comprises a plurality of post-processing memories, each one of said post-processing memories being disposed to store a pointer to a packet header.

19. A packet switch, comprising

a fetch stage coupled to a source of packet headers, said fetch stage being disposed to fetch at least portions of packet headers and present said portions of packet headers in parallel to a switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage;

said post-process stage coupled to said switching stage, and being disposed to perform protocol-specific processing on said packet headers, wherein said post-process stage comprises a post-processor coupled to said switching stage and disposed to alter at least a portion of a packet header responsive to a switching protocol.

20. A system, comprising

a packet memory;

a plurality of packet switches coupled to said packet memory;

a plurality of reorder memories coupled to said plurality of packet switches; and

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received;

wherein, each packet switch comprises a fetch stage coupled to said packet memory, said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage; and said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

21. A system as in claim 20, wherein said switching stage comprises a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results such that the in-out order of said packet header is preserved.

22. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory, said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of

14

packet headers in parallel to a switching stage, wherein said fetch stage comprises a fetch engine, said fetch engine being disposed to fetch a first block of M bytes of a packet header in response to a first signal, and being disposed to fetch an additional block of L bytes of said packet header in response to a second signal; and

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

23. A system as in claim 22, wherein said first signal indicates an empty fetch cache and wherein said second signal indicates a fetch cache with fewer than a selected number of unread bytes.

24. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage, wherein said fetch stage comprises a plurality of fetch caches, each one of said fetch caches being coupled to said source of packet headers and each being disposed to store at least a portion of a packet header; and

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

25. A system as in claim 24, wherein said switching stage comprises a plurality of switch engines, each being coupled to one said fetch cache and each being disposed to receive a packet header and to produce a set of results for switching said packet header.

26. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage, wherein said fetch stage comprises

a fetch engine coupled to said source of packet headers; and

a plurality of fetch caches coupled to said fetch engine, each said fetch cache comprising a plurality of buffers;

15

wherein said fetch engine is disposed to write at least a portion of each said packet header in sequence from each said fetch cache in a selected buffer thereof;

wherein said switching stage comprises a switch engine for each said fetch cache, wherein each said switch engine is disposed to read at least a portion of said packet header in sequence from each said buffer of said fetch cache;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

27. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage;

wherein said switching stage comprises

a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results for switching said packet header;

a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a packet header; and

a reorder/rewrite processor coupled to said plurality of reorder/rewrite memories and disposed to receive said packet headers from said reorder/rewrite memories in an order in which said packet headers were originally received;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

28. A system as in claim 27, wherein said reorder/rewrite processor is disposed to alter at least portions of said packet headers referenced by said reorder/rewrite memories.

29. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory, said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage;

wherein said switching stage comprises

a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results for switching said packet header; and

16

a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a packet header;

said reorder/rewrite memories being divided into sets, each said set of reorder/rewrite memories being assigned to and receiving outputs from exactly one said switch engine; and

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

30. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage;

wherein said switching stage comprises

a plurality of switching engines, each said switching engine having a plurality of reorder/rewrite memories coupled thereto, each said switching engine being disposed to write in sequence to one of said plurality of reorder/rewrite memories; and

a reorder/rewrite engine coupled to all said reorder/rewrite memories, said reorder/rewrite engine being disposed to read in sequence from said reorder/rewrite memories;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage.

31. A system, comprising

a packet memory;

a plurality of reorder memories;

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received; and

a plurality of packet switches coupled to said packet memory and said plurality of reorder memories, wherein each one of said plurality of packet switches comprises

a fetch stage coupled to said packet memory, said fetch stage being disposed to fetch packet headers from said packet memory and present at least portions of packet headers in parallel to a switching stage;

said switching stage coupled to said fetch stage, said switching stage being disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order to a post-process stage, wherein said switching stage comprises a plurality of reorder/rewrite memories, each one of said reorder/rewrite memories being disposed to store a pointer to a packet header.

17

32. A method of switching packets, said method comprising

fetching a sequence of packet headers corresponding to said packets from a source of said packet headers;
presenting said packet headers in parallel to a plurality of switch engines;

operating said switch engines to switch said packets asynchronously in parallel;

presenting switched packet headers in their original order to a post-processor; and

operating said post-processor to perform protocol-specific processing on said packet headers.

33. A method of switching packets, said method comprising

fetching a sequence of packet headers corresponding to said packets from a source of said packet headers, wherein said step of fetching includes fetching a first block of M bytes of a packet header in response to a first signal and fetching an additional block of L bytes of said packet header in response to a second signal;

presenting said packet headers in parallel to a plurality of switch engines;

operating said switch engines to switch said packets asynchronously in parallel;

presenting switched packet headers in their original order to a post-processor; and

operating said post-processor to perform protocol-specific processing on said packet headers.

34. A method as in claim 33, wherein said first signal indicates an empty fetch cache and wherein said second signal indicates a fetch cache with fewer than a selected number of unread bytes.

35. A method of switching packets, said method comprising

fetching a sequence of packet headers corresponding to said packets from a source of said packet headers, wherein said step of fetching includes storing said packet headers in sequence into a plurality of fetch caches;

presenting said packet headers in parallel to a plurality of switch engines;

operating said switch engines to switch said packets asynchronously in parallel;

presenting switched packet headers in their original order to a post-processor; and

operating said post-processor to perform protocol-specific processing on said packet headers.

36. A method of switching packets, said method comprising

fetching a sequence of packet headers corresponding to said packets from a source of said packet headers;

presenting said packet headers in parallel to a plurality of switch engines,

18

operating said switch engines to switch said packets asynchronously in parallel;

presenting switched packet headers in their original order to a post-processor; and

operating said post-processor to perform protocol-specific processing on said packet headers wherein said step of operating said post-processor stage comprises altering at least a portion of a packet header.

37. A method of switching packets, said method comprising

fetching a sequence of packet headers corresponding to said packets from a source of said packet headers;

presenting said packet headers in parallel to a plurality of switch engines;

operating said switch engines to switch said packets asynchronously in parallel, wherein said step of operating said switch engines includes coupling each said packet header to a selected fetch cache, coupling each said fetch cache to a selected switch engine, and coupling a set of results from said selected switch engine to a reorder/rewrite memory;

presenting switched packet headers in their original order to a post-processor; and

operating said post-processor to perform protocol-specific processing on said packet headers.

38. A method as in claim 37, wherein said reorder/rewrite memories are divided into sets, each said set of reorder/rewrite memories being assigned to and receiving outputs from exactly one said switch engine.

39. A system, including;

a packet memory;

a plurality of packet switches coupled to said packet memory, wherein each packet switch includes a fetch stage and a switching stage;

said fetch stage being coupled to said packet memory and disposed to fetch packet headers from said packet memory and present at least portions of said packet headers in parallel to said switching stage;

said switching stage being coupled to said fetch stage and disposed to switch said packet headers asynchronously in parallel and present said packet headers in their original order;

a plurality of reorder memories coupled to said plurality of packet switches; and

a reorder engine coupled to said plurality of reorder memories and disposed to receive packet headers from said reorder memories in an order in which they were originally received.

40. A system as in claim 39, wherein said switching stage comprises of a plurality of switch engines, each being disposed to receive a packet header and to produce a set of results such that the input order of said packet header is mimicked.

* * * * *